We thank the reviewers (RC1 & RC2) for their constructive comments, which have helped us substantially improve the manuscript. Please see details on changes made below. Reviewer comments are copied here (blue text), and associated responses follow each comment (black text), including in places copies of text inserted in the manuscript (black italics text).

Reviewer 1: RC1

The manuscript entitled "Harmonising and mapping Patagonian Shelf seabed sediment data" gathers about 500 surface sediment data to provide a reliable map of the Argentinean shelf. Worth noting that those types of mapping are crucial for several studies, such as habitat mapping and geohazards. Nevertheless, the authors make a major mistake when they treat the Patagonian shelf as a synonym for the Argentinean shelf. In fact, the Argentine shelf is much larger than the Patagonian area, and the manuscript must be corrected accordingly.

- In the revised text we now refer to the study region as the Argentine continental shelf, and state in the Introduction that this includes the Patagonian shelf as its southern sector. The manuscript is now titled 'Harmonising and mapping Argentine continental shelf seabed sediment data'.

The second aspect deals with Table 1. It is not clear which data were obtained from the samples in which the GSA method was classified as a "description". Indeed, most of the samples fall into this GSA method. In the case of Oliva (2020), the GSA method is "NA" (not available). Which data were used in this case?

- We now clarify in the caption of Table 1 that qualitative data/classification are based on sediment descriptions (e.g. core logs): *Table 1: Data sources for Argentine continental shelf seabed sediment study. Laser Particle Size Analyser (LPSA), Grain Size Analysis (GSA). Qualitative grain size data are based on sediment descriptions (GSA method = Description).* Where sampling or grain size methodology was not provided by the original author the table stated 'NA'. We have changed this to 'Not reported' in the revised manuscript to avoid confusion.

Concerning the kriging interpolation, I wonder which semivariogram model was used. I am not familiar with ArcGIS, and I do not know whether there is a basic model for determining the extent of a zone of influence for a given georeferenced value.

- RC1's final comment addresses the interpolation technique. Based on feedback from RC2 we have now revised the analysis and use indicator kriging, based on the large proportion of qualitative data and the fact that our Folk classes are categorical. In the methods sections we now state that experimental variograms were calculated for each indicator and fitted using spherical variogram models.

Reviewer 2: RC2

Roseby et al., bring together a variety of data sets collected between 1973 and 2023 from across the Patagonian shelf to facilitate simple mapping of the sediment type following the Folk class. The combined dataset is a useful resource and has the potential to be the foundation of future studies.

Overall, the manuscript is well written, and the figures are of high quality. The methodological approach to gather to compile the data is solid and well documented. However greater clarity is required concerning the grain size analysis:

Folk classification description could be clearer - The modified Folk-5 scheme differs from previous applications because mud >90% is isolated as a separate category. The rationale is reasonable, but the exact thresholds should be listed in a table in the main text or Sup Mat rather than relying mainly on Figure 2.

- In the caption for Figure 2, where we illustrate the different sediment categories that make up the Folk-7 and Folk-5 schemes, we now refer to supplementary materials, where the grain size thresholds for the different classes are defined.

Approximately 78% (380 of 486) of observations are qualitative descriptions that were translated into Folk classes using expert judgement. Although two researchers independently classified samples, there is no confusion matrix or uncertainty estimate for translation of descriptions and no sensitivity analysis showing how classification choices affect mapped outputs. Given that most data are qualitative, this uncertainty may be more important than interpolation uncertainty.

- We have reviewed the original comparison dataset and assessed agreement between the two independent researchers using the *caret* package in R (Kuhn, 2008). Agreement was assessed based on overall percentage agreement and Cohen's kappa. The following text is now included in the results section: *Regarding the classification of qualitative descriptions, when using the Folk-7 scheme, the independent researchers disagreed with 35 out of the total 482 surface samples, giving an overall agreement of 92.7% (95% CI: 90 – 95%). The most common mismatch was between Sand and muddy Sand (20/35; Supp. Figure 1). Cohen's Kappa (κ) for grain size classification using the Folk-7 scheme is 0.90, indicating near perfect agreement between the two researchers. When using the Folk-5 scheme the researchers disagreed with 11 out of the total 482 samples, giving an overall agreement of 97.7% (95% CI: 96-99%). Cohen's κ for grain size classification using the Folk-5 scheme is 0.96, again indicating near perfect agreement between the two researchers. The reduced number of samples that were mismatched when using the Folk-5 scheme is primarily due to the Folk-7 'Sand' and 'muddy Sand' grain size categories being combined (Fig. 2; Supp. Figure 2). In order to produce a final dataset that could be used to produce grain size maps for the shelf, minor differences in classification were discussed and resolved through consensus.*
- In the supplementary materials we have now included a confusion matrix for each modified Folk scheme which is referred to in the text (Supp. Figure 1 and 2).

The authors note that shell-hash samples were excluded from Folk coding while many shelly sediments were assigned to sand classes. This decision may affects comparisons with previous maps and has the potential to bias sediment proportions towards sands. This approach requires stronger justification and clearer documentation.

- In our methods setting we now provide justification for our decision to incorporate shells into our Folk Scheme, and not to create a separate category: *Previous maps of the Argentine continental shelf identified areas of the seafloor that are categorised as shell hash (Violante et al., 2014). Here we opted to follow Smeaton et al. (2021) by incorporating shelly sediments into Folk grain size categories, where they make up a component of the sediment (e.g. 'shelly muddy sand' is classed as 'muddy Sand'). Whilst this means that there will ultimately be differences between the maps presented here and those published previously, we made this decision for two reasons. Firstly, the presence of shells is typically seen in grain size descriptions (qualitative results), but quantitative data based on particle size does not discriminate between minerogenic and shell/biogenic material. Secondly, this dataset expands on harmonised grain* *size maps produced for other regions, and consistency would help in the combination and/or comparison of these (e.g. for the estimation of shelf carbon stocks; Smeaton et al., 2021).*

The interpolation of the data using Kriging approaches is simple, but I believe there are some issues that need clarified.

The authors assign numerical codes to sediment classes and then perform Ordinary Kriging on those codes before reclassifying the output. This is problematic as Folk classes are ordinal/categorical, not continuous numerical variables. Kriging assumes numerical continuity and meaningful distances between values. Alternative approaches such as indicator kriging, multinomial classification models, or interpolation on the underlying %gravel-%sand-%mud fractions would produce more robust mapping.

- We thank the reviewer for this suggestion, which we have implemented in the revised work. We now apply indicator kriging to predict the spatial probability of the different sediment classes that make up the Folk-5 and Folk-7 schemes. For each Folk class, indicator kriging is used to predict its spatial probability distribution. The final predicted Folk class for each grid cell is assigned as the class with the highest probability. These outputs also allow us to map the maximum probability assigned to each grid cell, highlighting areas with high and low probability/confidence. This has led to modifications of both the methods and results section; however, the key results of the study have not changed.

There is a lack of validation data. Model evaluation appears based on ArcGIS interpolation diagnostics (RMSE, Mean Error, R^2) these test the quality of model but not how it replicates the real world. Good model statistics does not equal a good map. The use of independent technique such as K-fold cross-validation or hold-out testing would provide greater confidence in the maps.

- We now run 10-fold cross-validation to estimate the performance of the indicator kriging model. The input data (surface samples) are randomly split into 10 folds. The indicator kriging model was then fitted 10 times, each time leaving out one fold as a validation dataset. After all folds had been evaluated, the predicted sediment classes are compared with the observed sediment classes for all validation samples using a confusion matrix. This is outlined in the methods section of the revised text. In order to carry out cross-validation in an efficient manner, the workflow was moved from ArcGIS to R and was supported by a colleague who has now been added as a co-author.

Missing details on the variogram. Ordinary Kriging results depend strongly on - Variogram model type, nugget, sill, range parameters. These are not reported in the manuscript therefore reproducing the interpolation therefore appears difficult.

- In the revised methods section, we now state that '*Experimental variograms were calculated independently for each indicator variable and fitted using spherical variogram models. The maximum number of neighbouring observations (nmax) was selected using 10-fold cross-validation (see below), with values of 10 and 20 providing the best predictive performance for the Folk-7 and Folk-5 datasets, respectively. Ordinary indicator kriging was subsequently performed for each sediment class using the selected nmax values to generate spatial predictions across the shelf. Indicator kriging was performed using the 'gstat' package in R (Pebesma, 2004), using a prediction grid with a cell size of 5 × 5 km (25 km²). The fitted variogram parameters (model type, nugget, sill, and range) are provided in Supplementary Tables 3 and 4.*'

The manuscript states that Ordinary Kriging was the “most appropriate” method. However, Table 2 shows very small differences between interpolation approaches, and Simple Kriging sometimes performs similarly or slightly better in some metrics. The wording should be softened.

- RC2’s final comment regarding the selection of ordinary kriging has now been resolved with the implementation of indicator kriging.